Machine Learning Methods for the Inversion of Hyperspectral Images

Mathieu Fauvel, Stéphane Girard, Sylvain Douté, Laurent Gardes

Chapter 1

MACHINE LEARNING METHODS FOR THE INVERSION OF HYPERSPECTRAL IMAGES

M. Fauvel[1], *S. Girard*[2,*], *S. Douté*[3], *and L. Gardes*[4]
[1] DYNAFOR, INRA & Université de Toulouse, France.
[2] Inria Grenoble Rhône-Alpes & Laboratoire Jean Kuntzmann, France.
[3] Institut de Planétologie et d'Astrophysique de Grenoble, CNRS-UGA, France,
[4] IRMA, Université de Strasbourg & CNRS, France.

Abstract

In this chapter, the physical analysis of planetary hyperspectral images by massive inversion is addressed. A direct radiative transfer model that relates a given combination of atmospheric or surface parameters to a spectrum is used to build a training set of synthetic observables. The inversion is based on the statistical estimation of the functional relationship between parameters and spectra. To deal with high dimensionality (image cubes typically present hundreds of bands), a two step method is proposed, namely K-GRSIR. It consists of a dimension reduction step followed by a regression with a non-linear least-squares algorithm. The dimension reduction is performed with the Gaussian Regularized Sliced Inverse Regression algorithm, which finds the most relevant directions in the space of synthetic spectra for the regression. The method is compared to several algorithms: a regularized version of k-nearest neighbors, partial least-squares, linear and non-linear support vector machines. Experimental results on simulated data sets have shown that non-linear support vector machines is the most accurate method followed by K-GRSIR. However, when dealing with real data sets, K-GRSIR gives the most interpretable results and is easier to train.

Key Words: Planetary hyperspectral images, Mars surface, regularized sliced inversion regression, support vector machines regression, kernel least-squares.

AMS Subject Classification: 62G08, 47A52, 85A25.

*E-mail address: Stephane.Girard@inria.fr

1. Introduction

For two decades, imaging spectroscopy has been a key technique for exploring planets [Murchie et al., 2007, Bibring et al., 2004b, Brown et al., 2004, Carlson et al., 1992]. Acquisition of several hundred thousands continuous spectra allows a fine characterization of the physical properties of the scene: Detection, mapping and characterization of minerals, as well as volatile species, whose presence often provide clues for the resolution of key climatic and geological issues. For instance, the OMEGA sensor acquires the spectral radiance coming from the planet in more than 200 contiguous spectral channels. A pixel of such an image is represented by a spectrum/vector $\mathbf{x} \in \mathbb{R}^d$, each component corresponds to a particular wavelength, d being the total number of wavelengths. Chemical composition, granularity, texture, and physical state are some of the parameters that characterize the morphology of spectra and thus the area of interest.

Deducing the physical parameters y from the observed spectra $\mathbf{x}$ is a central problem in geophysics, called an *inverse problem*. Since it generally cannot be solved analytically, optimization or statistical methods are necessary. Solving inverse problems requires an adequate understanding of the physics of the signal formation, *i.e.* a functional relation between $\mathbf{x}$ and y must be specified: $\mathbf{x} = g(y)$. Given g, different methods can be used to deduce the parameters from new observations. Current solutions to inverse problems can be divided into three main categories [Kimes et al., 2000]:

1. *Optimization algorithms*: These methods minimize an objective quality function that measures the fit between $\mathbf{x}$ and $g(y)$. Inverse problems are often ill-posed, therefore estimations can be unstable and a regularization is needed. For instance, a prior distribution on model parameters can be used. These approaches are computationally expensive since they independently invert new spectra. Therefore, they cannot be used to invert an image with several hundred thousand pixels. Moreover, they can sometimes fall into local minimum if the objective function is not convex.

2. *Look-up table (LUT) / k-nearest neighbors approaches (k-NN)*: A large database (LUT) is generated by a physical model for many combinations of parameter values. Each observed spectrum is then compared with the LUT spectra in order to find the best match (the nearest neighbor), according to an objective function minimization, typically the L_2 norm. Parameters are then deduced from this best match. The speed gain is significant in comparison to traditional optimization methods, since retrieving a value from memory is often faster than undergoing an expensive computation. The main disadvantages of this approach are the multiplicity of solutions and their instability.

3. *Training approaches*: A functional relation $y = f(\mathbf{x})$ between spectra and parameters is assumed, such as $f^{-1} \approx g$, and a LUT is used as a training set to estimate f. The advantage of such a training approach is that, once the relationship has been established, it can be used for very large sets and for all new images with the same physical model. Among training approaches, neural networks [Hastie et al., 2003, chapter 11] or support vector machines (SVM) [Hastie et al., 2003, chapter 12] seem promising but the underlying learning process remains time consuming [Combal et al., 2002, Durbha et al., 2007, Pragnère et al., 1999].

All approaches share the same difficulty when dealing with the increase of dimensionality: Parametric estimations are difficult [Hughes, 1968], distances between samples tend to be equal [Francois et al., 2007, Beyer et al., 1999] and the number of training samples needed for the training approaches becomes too high in practical situation [Fukunaga, 1990, chapter 5]. These problems are related to the *curse of dimensionality* [Donoho, 2000]. They make inverse problems in high dimension even more difficult and therefore algorithms must be robust to the dimension of the data.

One additional difficulty associated to planetary data is the very limited availability of ground truth or ground measurements to validate both the physical model and the training process. This motivates strategies to (i) assess the degree of adequacy between synthetic and real samples (ii) check if the estimated functional f - learned on the simulated data set - is still appropriate when inverting real images. As it will be seen in the experimental section, this assumption is not always fulfilled.

Several approaches are presented in this chapter to estimate the functional f: The well-known Support Vector Machines regression (SVM-R) [Hastie et al., 2003, Chapter 12], which works in full dimension, the Gaussian Regularized Sliced Inverse Regression (GRSIR) [Bernard-Michel et al., 2009d], which reduces the dimension before estimation, the Partial Least-Squares (PLS) regression [Hastie et al., 2003, Chapter 3] and the k-NN [Hastie et al., 2003, Chapter 13] still used by astrophysicists to perform massive inversion of hyperspectral images [Douté et al., 2001, Carlson et al., 2005]. Methods not relying on statistical models (SVM or k-NN) have two main advantages over parametric ones (GRSIR or PLS): No prior information is needed and no parameters estimation is necessary. However, in general, results are hardly interpretable and thus no information about the relationship between the input and the output is available. On the contrary, parametric methods reduce the dimension of spectra and the resulting subspace provides some physical information which can be used by astrophysicists [Bernard-Michel et al., 2009b]. Moreover, the training time is generally favorable to the parametric methods. In order to take advantage in a single algorithm of the learning ability of advanced non-parametric algorithms and of the interpretability of parametric algorithms, an extended version of GRSIR algorithm is proposed. First the dimension is reduced using GRSIR algorithm and then kernel least-squares (KLS) is used to learn the functional between reduced spectra and physical parameters. Motivations are the possibility to interpret the functional relationship in the reduced subspace and to reduce the training time while keeping the accuracy high.

Experiments are conducted on real hyperspectral images acquired from orbit of planet Mars by the OMEGA sensor [Bibring et al., 2004b]. In order to generate the simulated data, two physical models of solar light reflection by the surface have been considered, each one corresponding to a different geographical area of the planet.

Section 2 presents the different methods. Emphasis is put on SVM, GRSIR and the proposed K-GRSIR algorithm. The data sets are presented in Section 3 and experimental results are discussed in Section 4.

2. Inversion methods

For each method, a training set $\left(\mathbf{x}_i, y_i\right)_{i=1}^{n} \in \mathbb{R}^d \times \mathbb{R}^q$ is given and we try to estimate $f : y = f(\mathbf{x})$. Focussing on $\mathbb{R}$-valued functions, q functions are necessary to deduce q parameters

y of the physical model from the spectra x. Details about the simulation of the training and validation samples are provided in Section 3.

2.1. Support Vector Machines Regression

SVM are supervised methods for regression or estimation stemming from the machine learning theory. For inversion problems, the algorithm, which is called the ε-SVR, approximates the functional using solutions of the form

$$f(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i k(\mathbf{x}, \mathbf{x}_i) + b \tag{1}$$

where k is a kernel function and $\left((\alpha_i)_{i=1,\ldots,n},\ b\right)$ are the parameters of f which are estimated during the training process. The kernel k is used to produce non-linear functions. Given a training set, the training of an ε-SVR entails the following optimization problem:

$$\min \left[\frac{1}{n} \sum_{i=1}^{n} l\big(f(\mathbf{x}_i), y_i\big) + \lambda \|f\|^2 \right] \tag{2}$$

$$\text{with } l\big(f(\mathbf{x}), y\big) = \begin{cases} 0 \text{ if } |f(\mathbf{x}) - y| \leq \varepsilon \\ |f(\mathbf{x}) - y| - \varepsilon \text{ otherwise.} \end{cases}$$

This optimization problem is solved using the method of Lagrange multipliers [Vapnik, 1998, chapter 10]. The ε-SVR satisfies the sparsity constraint: Only some α_i are non-null which corresponding samples $\mathbf{x}_i$ are called "*Support Vectors*" (SVs). Some limitations come from the learning step involving a quadratic optimization. With a large training set, the training time can be very long. Moreover, the problem is exacerbated when several optimizations for parameter selection are considered. Despite of some recent works on quadratic solvers [Bottou et al., 2007], for large data set processing time remains large. Advanced methods can be used to select the optimal kernel parameters in an automatic procedure which reduces the processing time but requires more complex algorithmic tools [Moser and Serpico, 2009].

The choice of the kernel function is a crucial step with ε-SVR. A kernel function is a similarity measure between two samples and corresponds to a dot product in some feature space. To be an acceptable kernel, the function should be positive semi-definite [Camps-Valls and Bruzzone, 2009, chapter 2]. In a previous work [Bernard-Michel et al., 2009a], several kernels were investigated. It was found that Gaussian kernel provides the best results both in terms of accuracy and processing time:

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(- \frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2} \right). \tag{3}$$

This kernel is used for the experiments along with a linear kernel, *i.e* an inner product, the latter serves a basis of linear estimators.

Before running the algorithm, some hyperparameters need to be fitted:

- ε controls the resolution of the estimation. Large values produce rough approximations while small values produce fine estimations. It can be set using some prior on the signal to noise ratio.

- λ controls the smoothness of the solution. Large values imply nearly linear functions.
- σ is the Gaussian kernel parameter.

2.2. Gaussian Regularized Sliced Inverse Regression

To circumvent the "curse of dimensionality" effects, an alternative approach is to reduce the dimension of the data before the estimation. This is done by mapping the data onto a lower dimensional space and then doing the estimation:

$$y = f\big(\boldsymbol{\beta}^t \mathbf{x}\big), \tag{4}$$

where $\boldsymbol{\beta}^t\mathbf{x}$ denotes the projection on the subspace spanned by $\boldsymbol{\beta}$. In the following, the dimension of the projection subspace is denoted by $p < d$ whatever the dimension reduction method is.

The Principal Component Analysis (PCA) is surely one of the most used approach: $\boldsymbol{\beta}$ corresponds to the p first eigenvectors of the covariance matrix Σ of $\mathbf{x}$. Vectors $\boldsymbol{\beta}_\ell$ ($\ell = 1,\ldots,p$) maximize the variance of the projected components under unitary and orthogonality constraints on $\boldsymbol{\beta}_\ell$:

$$\hat{\boldsymbol{\beta}}_\ell^{pca} = \arg\max_{\boldsymbol{\beta}_\ell \in \mathbb{R}^d} \left[\boldsymbol{\beta}_\ell^t \hat{\Sigma} \boldsymbol{\beta}_\ell\right] \text{ with } \boldsymbol{\beta}_\ell^T \boldsymbol{\beta}_i = \delta_{\ell i},\ i = 1,\ldots,\ell \tag{5}$$

where $\hat{\Sigma}$ is the estimated covariance matrix and $\delta_{\ell i}$ is the Kronecker delta. It can be checked that they are the eigenvectors of $\hat{\Sigma}$.

However, in the case of a regression problem, PCA is generally not satisfying since only the explanatory variables $\mathbf{x}$ are considered and the dependent variable y is not taken into account. Specific feature extraction techniques have been developed for regression problems, and among them Sliced Inverse Regression (SIR) is very effective in high dimensional spaces [Li, 1991], see also [Girard and Saracco, 2014] for applications to astrophysics. The method consists of applying PCA to the inverse regression curve $\mathbb{E}(\mathbf{x}|y)$ (instead of applying it to the original predictor $\mathbf{x}$).

In the SIR methodology, $\mathbb{E}(\mathbf{x}|y)$ is estimated by a piecewise constant function thanks to a partitioning of the range of y into $h+1$ non-overlapping slices S_j (see Fig. 1). Under this approximation, SIR aims at maximizing the between slice variance under unitary variance and orthogonality constraints on the projected variables:

$$\hat{\boldsymbol{\beta}}_\ell^{sir} = \arg\max_{\boldsymbol{\beta}_\ell \in \mathbb{R}^d} \left[\boldsymbol{\beta}_\ell^t \hat{\Gamma} \boldsymbol{\beta}_\ell\right] \text{ with } \boldsymbol{\beta}_\ell^T \hat{\Sigma} \boldsymbol{\beta}_i = \delta_{\ell i},\ i = 1,\ldots,\ell \tag{6}$$

where $\hat{\Gamma}$ is the estimate of the covariance matrix of the inverse regression curve $\Gamma = \mathrm{cov}\big(\mathbb{E}(\mathbf{x}|y)\big)$:

$$\hat{\Gamma} = \sum_{j=1}^{H} \frac{n_j}{n} (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})^T,\ \bar{\mathbf{x}}_j = \frac{1}{n_j} \sum_{\mathbf{x}_i | y_i \in S_j} \mathbf{x}_i \tag{7}$$

with $\bar{\mathbf{x}}_j$ the between slice mean of slice S_j and $\bar{\mathbf{x}}$ the sample mean. Similarly to PCA, the projection axes are given by the eigenvectors corresponding to the largest p eigenvalues of $\hat{\Sigma}^{-1}\hat{\Gamma}$ [Li, 1991].

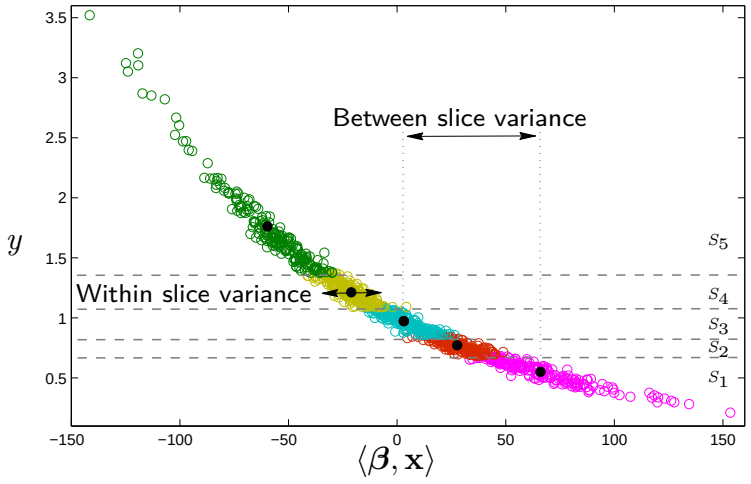


Figure 1. SIR: In this example, y is divided into 5 slices and the projection on $\boldsymbol{\beta}$ makes the estimation of the relationship $y = f(\boldsymbol{\beta}^t \mathbf{x})$ much easier ($\dim(\mathbf{x}) = 184$).

The quality of the projection is assessed with the SIR criterion (SIRC): It is the ratio of the between slice variance and the total variance:

$$\text{SIRC}(\boldsymbol{\beta}_\ell) = \frac{\boldsymbol{\beta}_\ell^T \hat{\Gamma} \boldsymbol{\beta}_\ell}{\boldsymbol{\beta}_\ell^T \hat{\Sigma} \boldsymbol{\beta}_\ell}, \tag{8}$$

with $0 \leq \text{SIRC} \leq 1$. From eq. (6), the higher the SIRC, the better the projection is. In high dimensional vector spaces, inverse problems are generally ill-posed [Tarantola, 2005, Aster et al., 2005], *i.e.* $\hat{\Sigma}$ is ill-conditioned making its inversion difficult. It thus has been proposed to compute a Gaussian Regularized version of Sliced Inverse Regression (GRSIR). Theoretical details can be found in [Bernard-Michel et al., 2009d]. The concept of this method is to incorporate some Gaussian prior on the projections in order to dampen the effect of noise in the input data, and to make the solution more regular or smooth. The ill-posed problem $\hat{\Sigma}^{-1}$ is then replaced by a slightly perturbed well-posed problem $\left(\Omega\hat{\Sigma} + \tau \mathbf{I}_d\right)^{-1}\Omega$, where τ is a positive regularization parameter, $\mathbf{I}_d$ is the identity matrix and Ω is a $d \times d$ matrix modeling the prior on the projection: It describes which directions are the most likely to contain $\boldsymbol{\beta}$ [Bernard-Michel et al., 2009d]. Finally, GRSIR consists of computing the eigenvectors corresponding to the largest eigenvalues of

$$\left(\Omega\hat{\Sigma} + \tau \mathbf{I}_d\right)^{-1}\Omega\hat{\Gamma}. \tag{9}$$

Using the eigenvalue decomposition of $\hat{\Sigma}$, several definitions of Ω have been proposed leading to several well known regularizations. Let us write

$$\hat{\Sigma} = \sum_{k=1}^{d} \delta_k \mathbf{v}_k \mathbf{v}_k^t \tag{10}$$

with $\delta_1 \geq \ldots \geq \delta_d$, the eigenvalues and $\mathbf{v}_k$ their associated eigenvectors. Then for all real valued function φ, Ω is defined as:

$$\Omega(\varphi) = \sum_{k=1}^{m} \varphi(\delta_k) \mathbf{v}_k \mathbf{v}_k^t \tag{11}$$

Table 1. SIR Regularization

$\varphi(\delta_k)$	m	Ω	Eigen problem	Regularization
$1/\delta_k$	$=d$	$\hat{\Sigma}^{-1}$	$\hat{\Sigma}^{-1}\hat{\Gamma}$	-
1	$=d$	$\mathbf{I}_d$	$\left(\hat{\Sigma}+\tau\mathbf{I}_d\right)^{-1}\hat{\Gamma}$	Ridge
1	$<d$	$\sum_{k=1}^{m}\mathbf{v}_k\mathbf{v}_k^t$	eq. (9)	PCA-Ridge
δ_k	$=d$	Σ	$\left(\hat{\Sigma}^2+\tau\mathbf{I}_d\right)^{-1}\hat{\Sigma}\hat{\Gamma}$	Tikhonov
δ_k	$<d$	$\sum_{k=1}^{m}\delta_k\mathbf{v}_k\mathbf{v}_k^t$	eq. (9)	PCA-Tikhonov

with $m \in \{1,\ldots,d\}$. Table 1 sums up the different proposed strategies. φ controls which directions of $\hat{\Sigma}$ that are favored: For instance, with the conventional SIR approach (first row of Table 1) directions corresponding to small variances are most likely, while no directions are privileged with ridge regularization. PCA based regularization approaches correspond to the situation where only directions with large variance are kept, *i.e.* a dimension reduction of $\hat{\Sigma}$ is done. For Tikhonov regularization, directions corresponding to large variances are most likely (in contrast to conventional SIR) but all directions are kept (in contrast to PCA based approaches). From previous works [Bernard-Michel et al., 2009a], the regularization is important but all the methods perform equally: no matter the method, it is just important to tune it correctly. In this chapter, ridge regularization is used because of its simplicity: only one parameter is to be tuned.

Once $\boldsymbol{\beta}$ is computed, a piecewise linear estimator is used, *i.e.* f in eq. (4) is a piecewise linear function (additional details can be found in [Bernard-Michel et al., 2009b]).

2.3. Partial Least-Squares regression

The PLS method is closely related to PCA (orthogonal projection onto lower dimensional space) and GRSIR (y is accounted for). PLS searches for the projection of the explanatory variable $\mathbf{x}$ onto a lower dimensional space that maximizes the covariance between $\mathbf{x}$ and y [Garthwaite, 1994]. It starts by computing the linear regression between $y_1 = y$ and $\mathbf{x}$ [Hastie et al., 2003, Chapter 3]:

$$\hat{\boldsymbol{\beta}}_1^{pls} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^d} \left[\operatorname{cov}^2\left(y_1, \langle\boldsymbol{\beta},\mathbf{x}\rangle\right)\right]. \tag{12}$$

Then the process is iterated on the residual y_2 between the prediction $\hat{y}_1$ and the true value y_1 ($y_2 = y_1 - \hat{y}_1$). Finally the $(\ell+1)$th PLS direction is found

$$\hat{\boldsymbol{\beta}}_\ell^{pls} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^d} \left[\operatorname{cov}^2\left(y_\ell, \langle\boldsymbol{\beta},\mathbf{x}\rangle\right)\right]. \tag{13}$$

The iteration stops when the number p of components is reached. The regression in the subspace is necessarily linear and may limit the efficiency of the method if the relationship between $\mathbf{x}$ and y is non linear.

2.4. k-Nearest Neighbors

The basic idea is to find from a LUT the k nearest spectra and to fix the estimated y as the mean parameter value of k nearest spectra parameter. The distance between two spectra is

taken as the Euclidean distance: $\|\mathbf{x}-\mathbf{x}_i\|^2$, where $\mathbf{x}$ is a spectrum from the image (to be inverted) and $\mathbf{x}_i$ a spectrum from the LUT. In the experiments, k is fixed to one, so only one neighbor is used for the estimation.

Although it has first been proposed by astrophysicists when performing massive inversion on hyperspectral images, k-NN is known to perform badly in high dimensional space. The reason is the *concentration of the measure* which makes all pairwise distances very similar [Francois et al., 2007, Beyer et al., 1999]. To dampen the effect of the dimensionality, the dimension of the spectra is reduced by PCA and then k-NN is applied on the reduced spectra

$$\tilde{\mathbf{x}} = \left[\frac{\mathbf{v}_1}{\sqrt{\delta_1}} \middle| \frac{\mathbf{v}_2}{\sqrt{\delta_2}} \middle| \dots \middle| \frac{\mathbf{v}_p}{\sqrt{\delta_p}} \right]^t \mathbf{x} \tag{14}$$

where $\mathbf{v}_i$ is the i^{th} first eigenvector of the covariance matrix $\hat{\Sigma}$ and δ_i its corresponding eigenvalue. The method is denoted by k-NN^{pca} in the following.

2.5. K-GRSIR

In the original GRSIR formulation [Bernard-Michel et al., 2009d], only the first axis is considered since it usually displays a very high SIRC (> 0.9). However, the following axes may have significant SIRC's. In this work, we thus propose to consider multiple axes instead of one, based on the SIRC value, and to use non-linear regression models similar to (1) fitted by a least-squares criteria.

Note that the original goal of GRSIR is not to reduce the dimensionality of the data but it is to find the central subspace (see section 2.2.). Contrary to PCA, it is not possible to link the eigenvalues and the reconstruction error of the data (because of the joint use of $\mathbf{x}$ and y). Therefore, the number p of selected eigenvectors cannot be chosen from the cumulative variance criteria. By looking at the scatter plot relating the projected data to the parameters, it appears that, for a too small SIRC, no signal can be seen and most of the variation comes from the noise, see Fig. 2. As a consequence, p is chosen by thresholding the SIRC, i.e. it is the number of eigenvectors whose SIRC is higher than 0.1 (10% of the total variance, see eq. (8)).

The regression function is similar to eq. (1), but a least-squares estimation is done, *i.e.* the ε-loss function of eq. (2) is changed to the quadratic loss:

$$\min \left[\frac{1}{n} \sum_{i=1}^{n} \left(f(\boldsymbol{\beta}^t \mathbf{x}_i) - y_i \right)^2 + \lambda \|f\|^2 \right]. \tag{15}$$

A close form solution of eq. (15) exits [Evgeniou et al., 2000]:

$$\boldsymbol{\alpha}_0 = (K_0 + \Lambda)^{-1} \mathbf{y}_0 \tag{16}$$

where $\boldsymbol{\alpha}_0$ is the vector of parameters, K_0 is the kernel matrix with the last row and last column completed with ones, the last element being set to zero, Λ is a diagonal matrix with its diagonal terms equal to λ except the last term which is null and $\mathbf{y}_0$ is the vector of dependent variables completed with one zero. This algorithm is called in the literature Kernel Least-Squares (KLS).

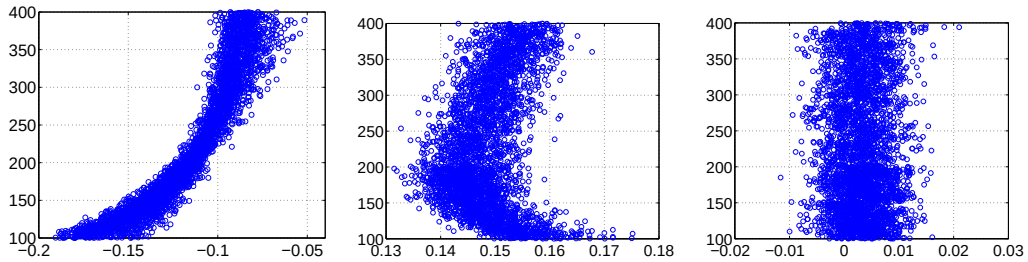


Figure 2. Scatter plot of the explanatory data projected on the central subspace direction versus dependent variable. The considered data are the simulated ones from the south polar permanent cap, and the physical parameter is the "Grain size of H_2O" ice, see Section 3 for details. Horizontal axis represents the projected data $\tilde{\mathbf{x}} = \beta_\ell^t \mathbf{x}, \ell \in \{1,2,5\}$ and vertical axis the physical parameter y.

3. Data sets

3.1. Mars Hyperspectral Images

In this chapter, a selection of images from the OMEGA sensor is analyzed. They pertain to two different types of martian terrains:

1. **South polar permanent cap (SPPC)**: Three OMEGA hyperspectral images are considered. They have been acquired during orbits 41, 61 and 103 that cover the high southern latitudes of Mars. The spatial resolution is about 2km per pixel and 184 wavelengths are considered in the range 0.95-4.15μm. For each image, a postprocessing aiming at correcting the atmospheric contribution in the spectra has been applied. See [Douté et al., 2007] for further details. These OMEGA observations revealed [Bibring et al., 2004a] that the south polar region of Mars mainly contains water ice, carbon dioxide ice and dust. A detailed qualitative mapping of H_2O and CO_2 ices during the local summer shows that the permanent south polar region is dominated by superficial CO_2 on the bright cap except at its edges where water ice appears in extended areas. Examining the coexistence modes (geographical or granular) between H_2O, CO_2 and dust that best explain the morphology of the spectra has led to a physical modeling of individual spectra with a surface reflectance model [Douté et al., 2007]. This model allows the generation of synthetic spectra with the corresponding sets of parameters that constitute a synthetic learning database, see paragraph 3.2. Here, we shall not work on the whole images to reverse the model because of the need of at least three different physical models required to simulate the whole image. We focus on the terrain unit characterized by a strong concentration of CO_2: The bright permanent south polar cap. This unit has been determined by a detection method based on wavelets [Schmidt et al., 2007]. For each image, the area dominated by CO_2 ice contains about 10000 to 20000 spectra.

2. **South polar seasonal deposits (SPSD)**: During southern winter, when the polar latitudes fall into darkness, temperature drops down to levels enabling intense conden-

sation of CO_2 in the solid form at the surface. Up to one third of the martian atmosphere can be trapped this way. During spring, when the sun returns, the energy balance of the icy deposits becomes positive and the CO_2 sublimates away, first at the lowest latitudes, and then closer and closer to the pole. OMEGA has acquired the most comprehensive set of observations to date in the near-infrared on the SPSD from mid-winter solstice to the end of the recession [Langevin et al., 2007]. Here we consider the "bright" part of the SPSD and the corresponding physical model derived from a previous study [Douté et al., 2008]: A substratum made of an intimate mixture of CO_2 and dust overlaid by a slab of CO_2 compact ice itself covered by a thin layer of dust contamination. We analyze an image acquired during orbit 1765 that covers the previous area.

3.2. Synthetic spectra (LUT)

From the above physical models, synthetic spectra have been generated with their corresponding physical parameters. Centered multidimensional Gaussian noise has been added, its covariance matrix was determined experimentally from the real images. For the validation sake, separate training and testing data sets have been randomly generated. The notations are: n (respectively n_t) is the number of spectra from the training data (respectively test data), $\mathbf{x}_i \in \mathbb{R}^d, i \in 1,\ldots,n$ denotes the spectra from the training data and $y_i \in \mathbb{R}, i \in 1,\ldots,n$ is one of the associated parameters (respectively $\check{\mathbf{x}}_j,\ \check{y}_j, j \in 1,\ldots,n_t$). Table 2 presents the number of samples for each data sets, the dimension of the data sets and the considered physical parameters.

4. Experiments

4.1. Synthetic spectra

In all experiments, parameters were selected by a 5-fold cross validation: Number of selected principal components for k-NN^{pca}, regularization parameters for (K-)GRSIR, kernel and regularization parameter for SVM or KLS and dimension of the subspace for PLS.

The quality of the estimation is assessed by computing the Normalized Root Mean Square Errors (NRMSE):

$$\text{NRMSE} = \sqrt{\frac{\frac{1}{n_t}\sum_{i=1}^{n_t}(\hat{y}_i - \check{y}_i)^2}{\frac{1}{n_t}\sum_{i=1}^{n_t}(\check{y}_i - \overline{y})^2}} \text{ with } \overline{y} = \frac{1}{n_t}\sum_{i=1}^{n_t}\check{y}_i \tag{17}$$

where $\check{y}_i$ is the real value and $\hat{y}_i$ the estimated one. It is the root mean square error normalized by the standard deviation of y. The indicator NRMSE is close to zero when the predicted values are accurate and becomes larger when predictions are poor. Results are reported in Table 3. Table 4 presents for some of the methods, the number of components kept. In what follows, a method per method analysis of the results is done.

Table 2. Simulated data characteristics.

	n	n_t	d	$p1$	$p2$	$p3$	$p4$	$p5$
SPPC	3528	3584	184	Prop. of H_2O	Prop. of CO_2	Prop. of dust	Grain size of H_20	Grain size of $C0_2$
SPSD	15250	15250	105	Prop. of CO_2	Prop. of dust	Prop. of H_2O	Ice width	Grain size of dust

Table 3. NRMSE for each data set and each method. LSVM stands for linear SVM, GSVM for Gaussian SVM. The two best results are emphasized.

	South polar permanent cap (SPPC)						
	k-NN	k-NN^{pca}	PLS	LSVM	GSVM	GRSIR	K-GRSIR
$p1$	1.32	0.26	0.34	0.34	*0.18*	0.28	*0.27*
$p2$	1.37	0.30	0.33	0.32	*0.18*	0.20	*0.19*
$p3$	1.45	0.27	0.25	0.24	*0.11*	0.16	*0.10*
$p4$	0.87	0.32	0.38	0.43	*0.26*	0.35	*0.28*
$p5$	0.26	0.37	0.31	0.27	*0.14*	0.18	*0.17*
Mean	1.05	0.30	0.31	0.32	*0.17*	0.23	*0.20*
	South polar seasonal deposits (SPSD)						
	k-NN	k-NN^{pca}	PLS	LSVM	GSVM	GRSIR	K-GRSIR
$p1$	0.94	0.28	0.24	0.24	*0.19*	0.22	*0.21*
$p2$	0.94	0.28	0.26	0.24	*0.19*	0.22	*0.21*
$p3$	1.02	0.87	0.86	0.81	*0.74*	0.96	*0.74*
$p4$	0.89	*0.17*	0.37	0.37	*0.16*	0.38	0.21
$p5$	1.15	0.41	0.56	0.48	*0.31*	0.41	*0.38*
Mean	0.98	0.40	0.46	0.42	*0.32*	0.43	*0.35*

Table 4. Number of components retained for each method for the different parameters.

	SPPC			**SPSD**		
	k-NN^{pca}	GRSIR	K-GRSIR	k-NN^{pca}	GRSIR	K-GRSIR
$p1$	4	1	3	3	1	1
$p2$	5	1	4	3	1	1
$p3$	5	1	6	28	1	3
$p4$	3	1	4	3	1	11
$p5$	4	1	4	3	1	5

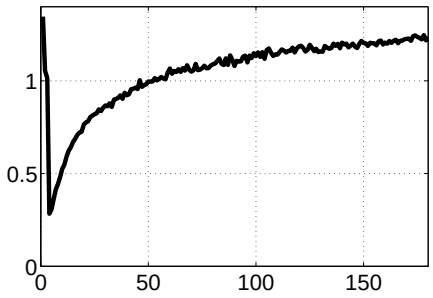


Figure 3. k-NN^{pca}. Cross validation errors as a function of dimension of the reduced spectra for the proportion of H_2O for the first model. Horizontally: The number of selected components. Vertically: The cross validation errors.

4.1.1. k-NN and k-NN^{pca}

In terms of NRMSE, k-NN^{pca} provides better results than its non-regularized version. The dimension of the subspace for each parameter is small, excepted for one that is part of the second physical model. Figure 3 shows the cross validation error as a function of the number of components selected for the proportion of H_2O. Using too few components increases sharply the NRMSE, while using more components progressively degrades the NRMSE. For the first data set, 3 components corresponds to 94.9% of the total variance, 4 components to 95.7% and 5 components to 96.2%. For the second data sets, 3 components corresponds to 98.6% and 28 to 99.4%.

4.1.2. PLS and linear SVM

Both methods seek for a linear functional and they give similar results. This is somewhat surprising though since the regression functions found by the two algorithms are different. Standard optimization for linear SVM makes it computationally too extensive in contrast to PLS. Compared to k-NN^{pca}, results are similar and no clear winner can be appointed, the rank of the method depending on the considered parameter.

4.1.3. GSVM

Non-linear SVM regression provides the best results in terms of NRMSE, for each parameter and for both models. The training time remains high, as it was discussed in [Bernard-Michel et al., 2009a]. This is especially true for the second simulated data sets, because a lot training samples are available.

As announced in the introduction, GSVM functional is hardly interpretable. However, the analysis of the support vectors indicates that saturation of some physical parameters exists in the model [Bernard-Michel et al., 2009c] (an increase of y does not change the spectra $\mathbf{x}$): The relation between the explanatory variable $\mathbf{x}$ and the dependent variable y is therefore highly non-linear and very hard to learn, for all algorithms.

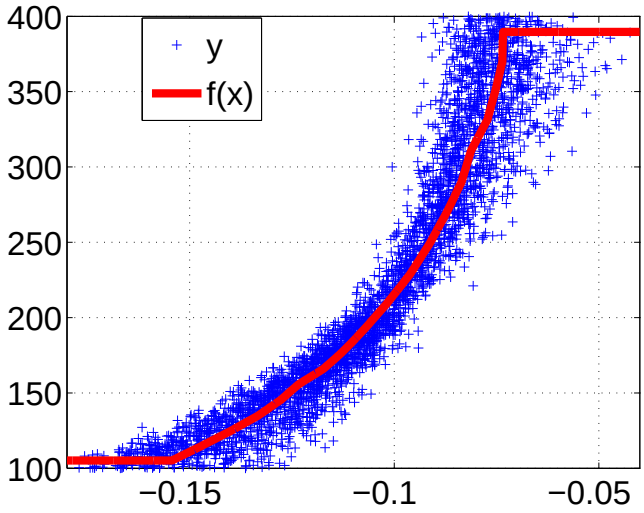


Figure 4. GRSIR relationship (in red) for the parameter grain size of H_2O ice for the first data set. Horizontal axis represents the projected data $\tilde{\mathbf{x}} = \beta^t \mathbf{x}$ and vertical axis the physical parameter y.

4.1.4. GRSIR

It provides better results in terms of NRMSE than linear or k-NN^{pca} methods for the first data set, and similar results for the second data sets. The higher number of training samples and the lower dimensionality of the second data set may explain why the difference in terms of NRMSE is smaller.

With GRSIR, it is possible to visualize the regression function eq. (4). Figure 4 presents the functional relationship found with GRSIR for the parameter "Grain size of H_2O" ice of the first model. It represents in red the piecewise linear function fitted with the projected data $\tilde{\mathbf{x}} = \beta^t \mathbf{x}$ and y. The corresponding SIRC is 0.87.

4.1.5. K-GRSIR

It performs slightly less accurately than GSVM in terms of NRMSE, and better than all other methods. Mean results are clearly better with K-GRSIR and get close to those of GSVM. Regarding the training time, the method performs in average 4.5 times faster than GSVM.

As a sanity check, we perform a comparison with the KLS learned on the first principal components, *i.e.* β are the p first components. Results on the first data sets demonstrate clearly that GRSIR is better suited for that regression problem. For parameter $p1$ the NRMSE is 0.67 for KLS against 0.27 for K-GRSIR. For parameter $p2$ it is 0.58 against 0.19. Similar results are obtained for the other parameters, confirming that for regression purpose, GRSIR is preferable to PCA.

Figure 5 presents the regression function found by K-GRSIR for the parameter "Grain size of H_2O" ice. The function is non-linear due to the use of a Gaussian kernel.

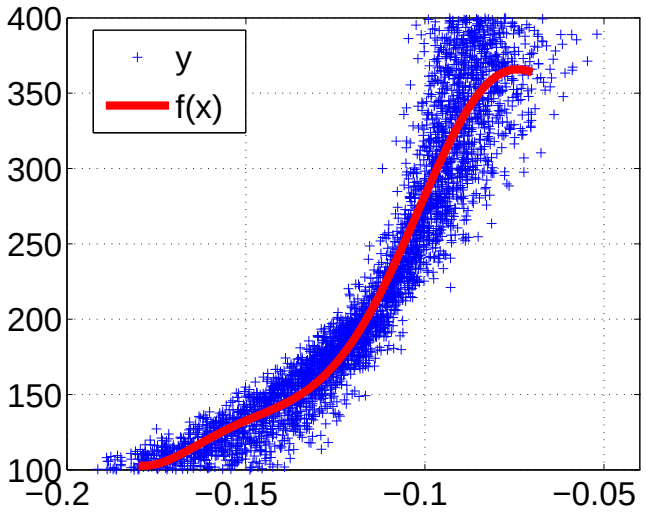


Figure 5. K-GRSIR relationship (in red) for the parameter grain size of H_2O ice for the first data set. Horizontal axis represents the projected data $\tilde{\mathbf{x}} = \beta^t \mathbf{x}$ and vertical axis the physical parameter y.

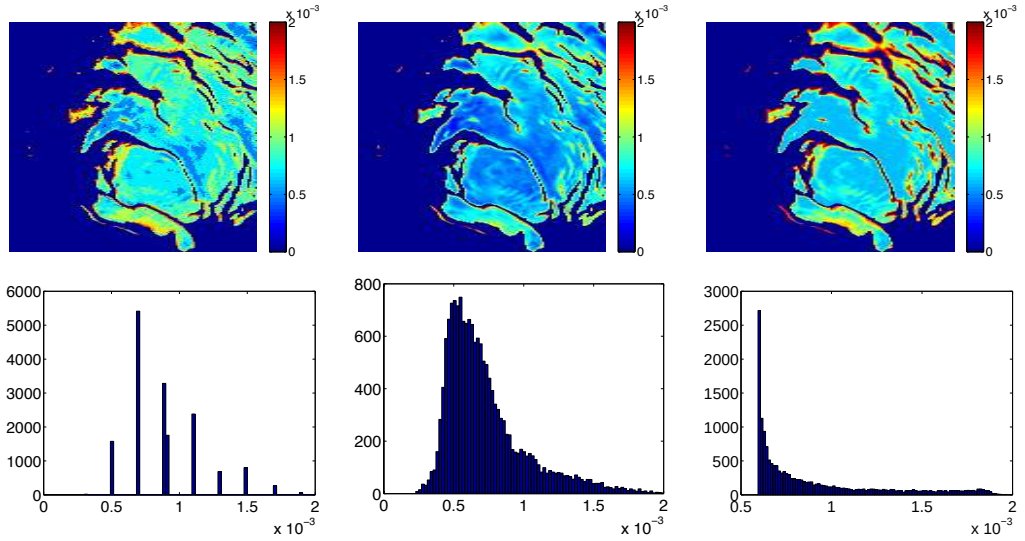
4.2. Mars Hyperspectral Images

In this section, the inversion of real OMEGA images is addressed. As mentioned in the introduction, no ground-truth is available. To assess the accuracy of inversion, we must rely on the *physical meaning* of the data: Proportions should be between 0 and 1, all estimations should be positive ... Another criteria is the stability of the estimation. For the south polar permanent cap, three images (orbit 41, 61 and 103) for the same geographical area were acquired in a short time span. No significant physical changes should occur, therefore the histograms of parameter values for the three inversions should be similar. In the following, real images corresponding to the SPPC are considered only.

Our original approach consists in learning the functional on the simulated data sets and then inverses the different images. However, from previous works [Bernard-Michel et al., 2009c, Bernard-Michel et al., 2009a], it was shown that this strategy failed to provide stable inversions. Small statistical variations between simulated and real data explain that problem [Bernard-Michel et al., 2009a]. Such variations affect the regularization parameters: They are chosen using simulated samples only and are not necessarily suited to the real images. For instance, for some badly constrained parameters, it makes the GRSIR axis almost orthogonal to the spectra of real images and thus spectral information is lost during the projection.

To ensure the stability of the retrieval, a range of values for a regularization parameter is tested to learn the functional, and to inverse the three images. The optimal value is chosen to minimize the variance between the mean values of the reconstructed physical parameters.

As an illustrative example, Fig. 6 focuses on the retrieval of the proportion of water by three different methods. It presents maps derived from observation 41 and their associated histograms. Note that inversion with GSVM was not performed, because the time processing was too important to be used in practical situation. Linear methods were not



(a) k-NN^{pca} (b) GRSIR (c) K-GRSIR

Figure 6. Proportion of water estimated by (a) k-NN^{pca}, (b) GRSIR, and (c) K-GRSIR from the hyperspectral image observed from orbit 41. Dark blue corresponds to pixels which are not considered for the inversion. The inverted image and the corresponding histogram are displayed for each method.

investigated after since they have performed poorly on the simulated data sets. Figure 7 presents results for the remaining parameters obtained with K-GRSIR.

A major drawback of the k-NN method against GRSIR and K-GRSIR can be seen on the histogram: It provides discrete estimation. Thus, the inversion map is non smooth and only values from the training set can be returned. GRSIR and, in particular K-GRSIR provides a smooth map. Some saturation exists for K-GRSIR on the left part of the histogram. The same saturation exists for third parameter, see Fig. 7(b), which is consistent with the constraint linking the first three parameters (the sum of compound abundances must equal one).

5. Conclusion

The physical analysis by massive inversion of hyperspectral images has been considered in this paper. Supervised parametric and non-parametric methods were investigated and compared to one commonly used method in astrophysics, k-NN. A method using both parametric and non-parametric algorithms has been proposed, namely K-GRSIR. On a simulated data set, this method performs slightly worse in terms of NRMSE than GSVM, but with the gain of a much reduced training time making it an accurate and tractable method in practical situations. On real images, GSVM is hard to train because of its computational load. K-GRSIR provides the most interpretable maps.

Considering Fig. 5, the regression curve seems to be polynomial, thus motivating the use of a polynomial kernel in K-GRSIR or in SVM. On simulated data, no difference

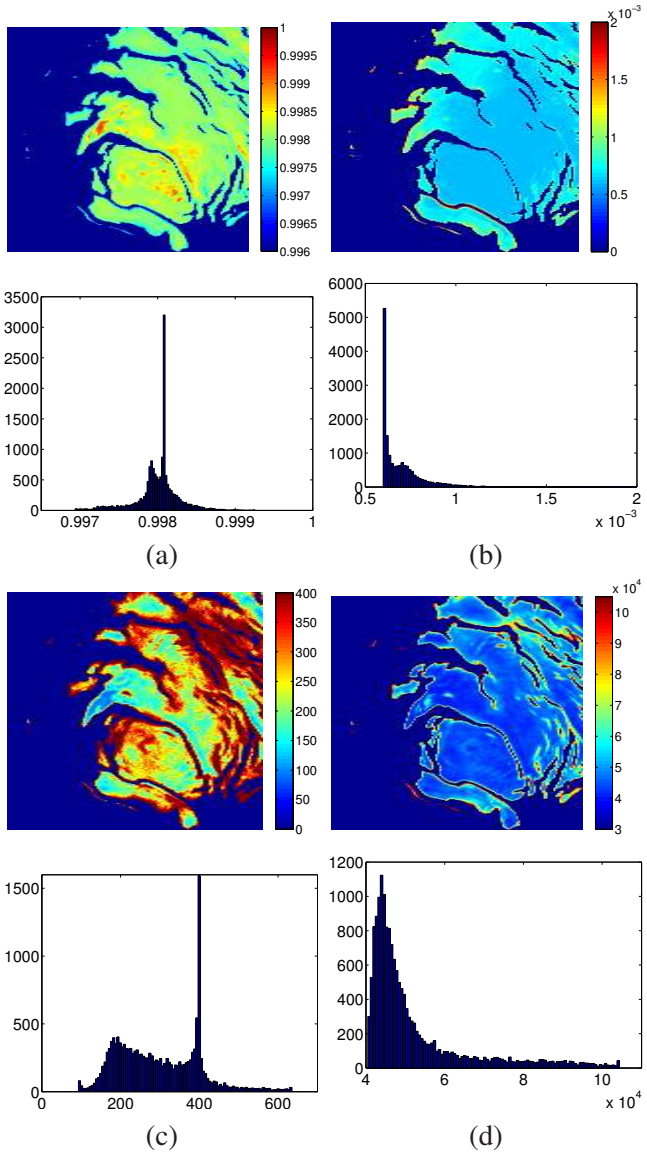


Figure 7. Orbit 41: (a) Proportion of CO_2, (b) Proportion of dust, (c) Grain size of H_2O and (d) Grain size of CO_2. The inversion is done with K-GRSIR. Dark blue corresponding to pixels which are not considered for the inversion. The inverted image and the corresponding histogram are displayed for each parameter.

in terms of NRMSE was found, but the training time is higher due to numerical problems [Bernard-Michel et al., 2009a]. The definition of a new kernel that would handle more efficiently the physical model is under investigation.

Current research concern the difference in terms of statistics between the simulated data and the real images. In particular, we are working on a semi-supervised framework to match statistics from simulated data and real data before the estimation of the regression function.

The use of multivariate versions of SIR [Coudret et al., 2014] able to deal with multidimensional response variable y would allow us to implement some constraints on the physical parameters (for instance, compound abundances must sum to one). This should improve the results obtained on real data sets. Also, the use of versions of SIR [Chiancone et al., 2016] robust with respect to outliers could be investigated.

Acknowledgement

This work is supported by a contract with CNES through its Groupe Système Solaire Program, by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01), by Inria and the Vahiné project (ANR-07-MDCO-013).

References

Aster, R., Borchers, B., and Thurber, C. (2005). *Parameter Estimation and Inverse Problems*. Elsevier Academic Press.

Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L., and Girard, S. (2009a). Machine learning techniques for the inversion of planetary hyperspectral images. In *Proc. of IEEE Int. Workshop on hyperspectral image and signal processing (WHISPERS-09)*, Grenoble.

Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L., and Girard, S. (2009b). Retrieval of mars surface physical properties from omega hyperspectral images using regularized sliced inverse regression. *Journal of Geophysical Research*, 114.

Bernard-Michel, C., Douté, S., Fauvel, M., Gardes, L., and Girard, S. (2009c). Support vectors machines regression for the estimation of Mars surface physical properties. In *European Symposium on Artificial Neural Networks, Advances in Computational Intelligence and Learning*.

Bernard-Michel, C., Gardes, L., and Girard, S. (2009d). Gaussian regularized sliced inverse regression. *Statistics and Computing*, 19:85–98.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful? In *In Int. Conf. on Database Theory*, pages 217–235.

Bibring, J.-P., Langevin, Y., Poulet, F., Gendrin, A., Gondet, B., Berthé, M., Soufflot, A., Drossart, P., Combes, M., G Bellucci, G., Moroz, V., Mangold, N., Schmitt, B., and the OMEGA team (2004a). Perennial water ice identified in the south polar cap of Mars. *Nature*, 428:627–630.

Bibring, J.-P., Soufflot, A., Berthé, M., Langevin, Y., Gondet, B., Drossart, P., Bouyé, M., Combes, M., Puget, P., Semery, A., Bellucci, G., Formisano, V., Moroz, V., Kottsov, V., Bonello, G., Erard, S., Forni, O., Gendrin, A., Manaud, N., Poulet, F., Poulleau, G., Encrenaz, T., Fouchet, T., Melchiori, R., Altieri, F., Ignatiev, N., Titov, D., Zasova, L., Coradini, A., Capacionni, F., Cerroni, P., Fonti, S., Mangold, N., Pinet, P., Schmitt, B., Sotin, C., Hauber, E., Hoffmann, H., Jaumann, R., Keller, U., Arvidson, R., Mustard, J., and F., F. (2004b). *OMEGA: Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité*, pages 37–49. ESA SP-1240: Mars Express: the Scientific Payload.

Bottou, L., Chapelle, O., DeCoste, D., and Weston, J., editors (2007). *Large Scale Kernel Machines*. MIT Press, Cambridge, MA.

Brown, R. H., Baines, K. H., Bellucci, G., Bibring, J.-P., Buratti, B. J., Capaccioni, F., Cerroni, P., Clark, R. N., Coradini, A., Cruikshank, D. P., Drossart, P., Formisano, V., Jaumann, R., Langevin, Y., Matson, D. L., McCord, T. B., Mennella, V., Miller, E., Nelson, R. M., Nicholson, P. D., Sicardy, B., and Sotin, C. (2004). The Cassini Visual And Infrared Mapping Spectrometer (Vims) Investigation. *Space Science Reviews*, 115:111–168.

Camps-Valls, G. and Bruzzone, L., editors (2009). *Kernel Methods for Remote Sensing Data Analysis*. Wiley.

Carlson, R., Anderson, M., Mehlman, R., and Johnson, R. (2005). Distribution of hydrate on Europa: Further evidence for sulfuric acid hydrate. *Icarus*, 177(2):461–471.

Carlson, R. W., Weissman, P. R., Smythe, W. D., Mahoney, J. C., the NIMS Science, and Teams, E. (1992). Near infrared spectrometer experiment on Galileo. *Space Science Reviews*, 60:457–502.

Chiancone, A., Forbes, F., and Girard, S. (2016). Student Sliced Inverse Regression. *Computational Statistics and Data Analysis*, to appear.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., and Wang, L. (2002). Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, 84:1–15.

Coudret, R., Girard, S., and Saracco, J. (2014). A new sliced inverse regression method for multivariate response. *Computational Statistics and Data Analysis*, 77:285–299.

Donoho, D. L. (2000). High-dimensional data analysis: the curses and blessing of dimensionality. In *AMS Mathematical challenges of the 21st century*.

Douté, S., Deforas, E., Schmidt, F., Oliva, R., and Schmitt, B. (2007). A Comprehensive Numerical Package for the Modeling of Mars Hyperspectral Images. In *Lunar and Planetary Institute Conference Abstracts*, volume 38 of *Lunar and Planetary Institute Conference Abstracts*, pages 1836–1837.

Douté, S., Schmidt, F., Schmitt, B., Langevin, Y., Vincendon, M., Bibring, J.-P., and Omega Team (2008). Physical Characterization of the South Seasonal Cap of Mars During Recession from OMEGA Observations. In *Lunar and Planetary Institute Science Conference Abstracts*, volume 39 of *Lunar and Planetary Institute Science Conference Abstracts*.

Douté, S., Schmitt, B., Bibring, J.-P., Langevin, Y., Altieri, F., Bellucci, G., Gondet, B., and the Mars Express OMEGA Team (2007). Nature and composition of the icy terrains from Mars express OMEGA observations. *Planetary and Space Science*, 55:113–133.

Douté, S., Schmitt, B., Lopes-Gautier, R. M. C., Carlson, R. W., Soderblom, L., and Shirley, J. (2001). Mapping SO_2 frost on Io by the modeling of NIMS hyperspectral images. *Icarus*, 149:107–132.

Durbha, S., King, R., and Younan, N. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, 107:348–361.

Evgeniou, T., Pontil, M., and Poggio, T. (2000). Regularization networks and support vector machines. *Advances on Computational Mathematics*.

Francois, D., Wertz, V., and Verleysen, M. (2007). The concentration of fractional distances. *IEEE Trans. Knowl. Data Eng.*, 19(7):873–886.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. CA: Academic Press, San Diego.

Garthwaite, P. H. (1994). An interpretation of partial least squares. *Journal of the American Statistical Association*, 89(425):122–127.

Girard, S. and J. Saracco, J. (2014). An introduction to dimension reduction in nonparametric kernel regression, In D. Fraix-Burnet and D. Valls-Gabaud, editors, *Regression methods for astrophysics*, volume 66, pages 167–196, EDP Sciences.

Hastie, T., Tibshirani, R., and Friedman, J. (2003). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hughes, G. F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory*, IT-14:55–63.

Kimes, D., Knyazikhin, Y., Privette, J., Abuegasim, A., and Gao, F. (2000). Inversion methods for physically-based models. *Remote Sensing Reviews*, 18:381–439.

Langevin, Y., Bibring, J.-P., Montmessin, F., Forget, F., Vincendon, M., Douté, S., Poulet, F., and Gondet, B. (2007). Observations of the south seasonal cap of Mars during recession in 2004-2006 by the OMEGA visible/near-infrared imaging spectrometer on board Mars Express. *Journal of Geophysical Research E: Planets*, 112(E8):E08S12.

Li, K. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86:316–327.

Moser, G. and Serpico, S. (2009). Automatic parameter optimization for support vector regression for land and sea surface temperature estimation from remote sensing data. *IEEE Trans. Geosci. Remote Sens.*, 47(3):909–921.

Murchie, S., Arvidson, R., Bedini, P., Beisser, K., Bibring, J.-P., Bishop, J., Boldt, J., Cavender, P., Choo, T., Clancy, R. T., Darlington, E. H., Des Marais, D., Espiritu, R., Fort, D., Green, R., Guinness, E., Hayes, J., Hash, C., Heffernan, K., Hemmler, J., Heyler, G., Humm, D., Hutcheson, J., Izenberg, N., Lee, R., Lees, J., Lohr, D., Malaret, E., Martin, T., McGovern, J. A., McGuire, P., Morris, R., Mustard, J., Pelkey, S., Rhodes, E., Robinson, M., Roush, T., Schaefer, E., Seagrave, G., Seelos, F., Silverglate, P., Slavney, S., Smith, M., Shyong, W.-J., Strohbehn, K., Taylor, H., Thompson, P., Tossman, B., Wirzburger, M., and Wolff, M. (2007). Compact reconnaissance imaging spectrometer for Mars (CRISM) on Mars reconnaissance orbiter (MRO). *Journal of Geophysical Research (Planets)*, 112(E11):E05S03.

Pragnère, A., Baret, F., Weiss, M., Myneni, R., Knyazikhin, Y., and Wang, L. (1999). Comparison of three radiative transfer model inversion techniques to estimate canopy biophysical variables from remote sensing data. *Geoscience and Remote Sensing Symposium, 1999. IGARSS '99 Proceedings. IEEE 1999 International*, 2:1093–1095.

Schmidt, F., Douté, S., and Schmitt, B. (2007). WAVANGLET: An efficient supervised classifier for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.*, 45(5):1374–1385.

Tarantola, A. (2005). *Inverse problem theory and model parameter estimation*. SIAM.

Vapnik, V. (1998). *Statistical Learning Theory*. Wiley, New York.